

04.25.01


09830244-042PG1

JC03 Rec'd PCT/PTO

23 APR 2001

FORAM 1390 U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE		ATTORNEY'S DOCKET NUMBER PF-0620 USN
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371		U.S. APPLICATION NO. TO BE ASSIGNED 09/830244
INTERNATIONAL APPLICATION NO. PCT/US99/24046	INTERNATIONAL FILING DATE 22 October 1999	PRIORITY DATE CLAIMED 23 October 1998
TITLE OF INVENTION LYSINE-RICH STATHERIN PROTEIN		
APPLICANT(S) FOR DO/EO/US INCYTE PHARMACEUTICALS, INC.; TANG, Y. Tom; CORLEY, Neil C.; GUEGLER, Karl J.; PATTERSON, Chandra		
Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information: 1. <input checked="" type="checkbox"/> This is the FIRST submission of items concerning a filing under 35 U.S.C. 371. 2. <input type="checkbox"/> This is a SECOND or SUBSEQUENT submission of items concerning a filing under 35 U.S.C. 371. 3. <input type="checkbox"/> This is an express request to promptly begin national examination procedures (35 U.S.C. 371 (f)). 4. <input type="checkbox"/> The US has been elected by the expiration of 19 months from the priority date (PCT Article 31). 5. <input checked="" type="checkbox"/> A copy of the International Application as filed (35 U.S.C. 371(c)(2)) a. <input type="checkbox"/> is attached hereto (required only if not communicated by the International Bureau) b. <input type="checkbox"/> has been communicated by the International Bureau. c. <input checked="" type="checkbox"/> is not required, as the application was filed in the United States Receiving Office (RO/US). 6. <input type="checkbox"/> An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)). 7. <input checked="" type="checkbox"/> Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3)) a. <input type="checkbox"/> are attached hereto (required only if not communicated by the International Bureau). b. <input type="checkbox"/> have been communicated by the International Bureau. c. <input type="checkbox"/> have not been made; however, the time limit for making such amendments has NOT expired. d. <input checked="" type="checkbox"/> have not been made and will not be made. 8. <input type="checkbox"/> An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)). 9. <input checked="" type="checkbox"/> An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)). 10. <input type="checkbox"/> An English language translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).		
Items 11 to 16 below concern document(s) or information included:		
11. <input type="checkbox"/> An Information Disclosure Statement under 37 CFR 1.97 and 1.98. 12. <input type="checkbox"/> An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.27 and 3.31 is included. 13. <input type="checkbox"/> A FIRST preliminary amendment. <input type="checkbox"/> A SECOND or SUBSEQUENT preliminary amendment. 14. <input type="checkbox"/> A substitute specification. 15. <input type="checkbox"/> A change of power of attorney and/or address letter. 16. <input checked="" type="checkbox"/> Other items or information: 1) Transmittal Letter (2 pp. in duplicate) 2) Return Postcard 3) Express Mail Label No.: EL 856 112 999 US 4) Request to Transfer		

JC18 Rec'd PCT/PTO 2 3 APR 2001

U.S. APPLICATION NO. (if known) see 37 CFR 1.5 TO BE A REPLY 09/830244		INTERNATIONAL APPLICATION NO.: PCT/US99/ 24046		ATTORNEY'S DOCKET NUMBER PF-0620 USN	
17. <input checked="" type="checkbox"/> The following fees are submitted: BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(5)): Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO.....\$1000.00 <input type="checkbox"/> International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO..\$860.00 International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO.....\$710.00 <input checked="" type="checkbox"/> International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provisions of PCT Article 33(1)-(4).....\$690.00 <input type="checkbox"/> International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4).....\$100.00					
ENTER APPROPRIATE BASIC FEE AMOUNT =				\$690.00	
Surcharge of \$130.00 for furnishing the oath or declaration later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(e)).				\$	
CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total Claims	20 =	0	X \$ 18.00	\$	
Independent Claims	2 =	0	X \$ 80.00	\$	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ \$270.00	\$	
TOTAL OF ABOVE CALCULATIONS =				\$690.00	
<input type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.				\$	
SUBTOTAL =				\$690.00	
Processing fee of \$130.00 for furnishing the English translation later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(f)).				\$	
TOTAL NATIONAL FEE =				\$690.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by the appropriate cover sheet (37 CFR 3.28, 3.31). \$40.00 per property +				\$	
TOTAL FEES ENCLOSED =				\$690.00	
				Amount to be Refunded:	\$
				Charged:	\$
a. <input type="checkbox"/> A check in the amount of \$ _____ to cover the above fees is enclosed. b. <input checked="" type="checkbox"/> Please charge my Deposit Account No. <u>09-0108</u> in the amount of \$ <u>690.00</u> to cover the above fees. c. <input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. <u>09-0108</u> . A duplicate copy of this sheet is enclosed.					
NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.					
SEND ALL CORRESPONDENCE TO:					
INCYTE GENOMICS, INC. 3160 Porter Drive Palo Alto, CA 94304			 SIGNATURE		
NAME: Diana Hamlet-Cox					
REGISTRATION NUMBER: 33,302					
DATE: <u>23</u> April 2001					

LYSINE-RICH STATHERIN PROTEIN

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of a lysine-rich statherin protein and to the use of these sequences in the diagnosis, treatment, and prevention of neoplastic and autoimmune/inflammatory disorders, and infectious diseases.

BACKGROUND OF THE INVENTION

Low molecular weight proteins, polypeptide chains that are less than about one hundred amino acid residues (aa), have become the focus of investigation as components of important molecular and biochemical pathways during cell growth, tissue differentiation, inflammation disorders, and the immune response. Examples of such proteins are histatins and statherins.

Human histatins are a family of low molecular weight (51 to 57 aa), neutral to very basic, histidine-rich salivary proteins. Histatins are believed to function as part of the nonimmune host defense system in the oral cavity. Salivary histatins (HSTs) are potent in vitro antifungal and antibacterial agents and have great promise as therapeutic agents in humans with oral candidiasis. Candida albicans is the predominant species of yeast isolated from patients with oral candidiasis, which is frequently a symptom of human immunodeficiency virus (HIV) infection and is a criterion for staging and progression of AIDS. ¹²⁵I-HST binding studies indicated that C. albicans expressed a class of saturable binding sites, numbering 8.6×10^5 sites/cell. HSTs are also effective in killing another medically important opportunistic fungal pathogen, Cryptococcus neoformans, which has become a new threat among immunocompromised patients, including those with AIDS (Tsai, H. and L.A. Bobek (1997) Biochim. Biophys. Acta 1336:367-369).

Human statherin (STATH) is a low-molecular weight (43 aa) acidic phosphoprotein that acts as an inhibitor of precipitation of calcium phosphate salts in the oral cavity. STATH also has affinity to hydroxyapatite and interacts with oral bacteria on absorption to hydroxyapatite (Lamkin, M.S. and F.G. Oppenheim (1993) Crit. Rev. Oral. Biol. Med. 4:251-259). The HST and STATH genes show nearly identical overall gene structures. The HST and STATH loci exhibit 77%-81% sequence identity in intron DNA and 80%-88% sequence identity in noncoding exons but only 38%-43% sequence identity in the protein-coding regions of exons 4 and 5. It has been suggested that HST and STATH belong to a single gene family exhibiting accelerated evolution between the HST and STATH coding sequences (Sabatini, L.M. et al. (1993) Mol. Biol. Evol. 10:497-511).

The discovery of a new lysine-rich statherin protein and the polynucleotides encoding it satisfies a need in the art by providing new compositions which are useful in the diagnosis,

prevention, and treatment of neoplastic and autoimmune/inflammatory disorders, and infectious diseases.

SUMMARY OF THE INVENTION

5 The invention features substantially purified polypeptides, lysine-rich statherin protein, referred to as "LRSP." In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof. The invention also includes a polypeptide comprising an amino acid sequence that differs by one or more conservative amino acid substitutions from an amino acid sequence selected
10 from the group consisting of SEQ ID NO:1.

The invention further provides a substantially purified variant having at least 90% amino acid identity to at least one of the amino acid sequences selected from the group consisting of SEQ ID NO:1 and fragments thereof. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of
15 SEQ ID NO:1 and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof.

Additionally, the invention provides an isolated and purified polynucleotide which hybridizes
20 under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof.

25 The invention also provides a method for detecting a polynucleotide in a sample containing nucleic acids, the method comprising the steps of: (a) hybridizing the complement of the polynucleotide sequence to at least one of the polynucleotides of the sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide in the sample. In one aspect,
30 the method further comprises amplifying the polynucleotide prior to hybridization.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:2 and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide sequence selected from the group consisting of

SEQ ID NO:2 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:2 and fragments thereof.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing a polynucleotide of the invention under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide selected from the group consisting of SEQ ID NO:1 and fragments thereof. The invention also provides a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of LRSP, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of LRSP, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof.

BRIEF DESCRIPTION OF THE FIGURES AND TABLE

Figures 1A, 1B, and 1C show the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ ID NO:2) of LRSP. The alignment was produced using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA).

Figures 2A, 2B, 2C, 2D, 2E, 2F, 2G, and 2H show the nucleotide sequence alignment between LRSP (Incyte Clone ID 2820214; SEQ ID NO:2), human statherin (GI 338507; SEQ ID NO:5), and human basic histidine-rich protein (GI 179465; SEQ ID NO:6), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

Figure 3 shows the amino acid sequence alignment between LRSP (Incyte Clone ID 2820214; SEQ ID NO:1), human statherin (GI 338508; SEQ ID NO:3), and human basic histidine-rich protein (GI 179466; SEQ ID NO:4), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

5 Table 1 shows the tools, programs, and algorithms used to analyze LRSP, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood
10 that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an,"
15 and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same
20 meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might
25 be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"LRSP" refers to the amino acid sequences of substantially purified LRSP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and
30 human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of LRSP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of LRSP either by directly interacting with LRSP or by acting on components of the biological pathway in which LRSP participates.

An "allelic variant" is an alternative form of the gene encoding LRSP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding LRSP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as LRSP or a polypeptide with at least one functional characteristic of LRSP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding LRSP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding LRSP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent LRSP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of LRSP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of LRSP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small

molecules, or any other compound or composition which modulates the activity of LRSP either by directly interacting with LRSP or by acting on components of the biological pathway in which LRSP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind LRSP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic LRSP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" and "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which

WO 00/24779

PCT/US99/24046

depend upon binding between nucleic acid strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding LRSP or fragments of LRSP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of one or more Incyte Clones and, in some cases, one or more public domain ESTs, using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that, when made, least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
25	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
30	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
35	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
40	Tyr	His, Phe, Trp

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "fragment" is a unique portion of LRSP or the polynucleotide encoding LRSP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:2 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:2, for example, as distinct from any other sequence in the same genome. A fragment of SEQ ID NO:2 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:2 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:2 and the region of SEQ ID NO:2 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1 is encoded by a fragment of SEQ ID NO:2. A fragment of SEQ ID NO:1 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1.

For example, a fragment of SEQ ID NO:1 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1. The precise length of a fragment of SEQ ID NO:1 and the region of SEQ ID NO:1 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

5 The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization
10 assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to
15 one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

20 The phrases “percent identity” and “% identity,” as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and “diagonals saved”=4. The “weighted” residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polynucleotide sequence pairs.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms

<http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis

programs including “blastn,” that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called “BLAST 2 Sequences” that is used for direct pairwise comparison of two nucleotide sequences. “BLAST 2 Sequences” can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The “BLAST 2 Sequences” tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases “percent identity” and “% identity,” as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some

WO 00/24779

PCT/US99/24046

alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the hydrophobicity and acidity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific

hybridization is an indication that two nucleic acid sequences share a high degree of identity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Generally, such wash temperatures are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate).

to which cells or their nucleic acids have been fixed).

The words “insertion” and “addition” refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

“Immune response” can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term “microarray” refers to an arrangement of distinct polynucleotides on a substrate.

The terms “element” and “array element” in a microarray context, refer to hybridizable
10 polynucleotides arranged on the surface of a substrate.

The term “modulate” refers to a change in the activity of LRSP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of LRSP.

The phrases “nucleic acid” and “nucleic acid sequence” refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably
20 linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

“Peptide nucleic acid” (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Probe" refers to nucleic acid sequences encoding LRSP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

“Primers” are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel et al., 1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

14

partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence.

5 This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter
10 sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

15 The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding LRSP, or fragments thereof, or LRSP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a
20 protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A
25 and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

30 A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells,

trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

“Transformation” describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term “transformed” cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A “variant” of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an “allelic” (as defined above), “splice,” “species,” or “polymorphic” variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass “single nucleotide polymorphisms” (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A “variant” of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

WO 00/24779

PCT/US99/24046

The invention is based on the discovery of a new human lysine-rich statherin protein (LRSP), the polynucleotides encoding LRSP, and the use of these compositions for the diagnosis, treatment, or prevention of neoplastic and autoimmune/inflammatory disorders, and infectious diseases.

Nucleic acids encoding the LRSP of the present invention were identified in Incyte Clone 5 2820214 from the breast cDNA library (BRSTNOT14) using a computer search for nucleotide and/or amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 2820214H1 and 2820214T6 (BRSTNOT14), 3600320F6 (DRGTNOT01), 4254615H1 (BSCNNOT03), 2656117F6 (THYMNOT04), and 1325717F6 (LPARNOT02).

10 In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1, as shown in Figures 1A, 1B, and 1C. LRSP is 95 amino acids in length; has two potential N-glycosylation sites at residues N24 and N76, one potential casein kinase II phosphorylation site at residue T15, and one potential protein kinase C phosphorylation site at residue S29; and is rich in basic (lysine, arginine, and histidine) and aliphatic amino acid residues, containing 15 21% and 26% of these, respectively. As shown in Figures 2A, 2B, 2C, 2D, 2E, 2F, 2G, and 2H, the nucleotide sequence encoding LRSP (Incyte Clone ID 2820214; SEQ ID NO:2) has chemical similarity with the nucleotide sequence encoding human statherin (GI 338507; SEQ ID NO:5) and with the nucleotide sequence encoding human basic histidine-rich protein (GI 179465; SEQ ID NO:6). In particular, the region of the nucleotide sequence encoding LRSP from about nucleotide 20 1059 to about nucleotide 1331 is similar to a region of the nucleotide sequence of human statherin (94% identity) and LRSP is therefore considered a splice variant. In addition the region of the nucleotide sequence encoding LRSP from about nucleotide 1122 to about nucleotide 1331 is similar to a region of the nucleotide sequence of human basic histidine-rich protein (68% identity). As shown in Figure 3, LRSP (Incyte Clone ID 2820214; SEQ ID NO:1) has chemical and structural similarity 25 with human statherin (GI 338508; SEQ ID NO:3) and human basic histidine-rich protein (GI 179466; SEQ ID NO:4). In particular, LRSP and human statherin share 14% identity, and conserved residues at F3, D20, P45, P48, and P61 in LRSP. In addition, LRSP and human basic histidine-rich protein have similar isoelectric points, 10.5 and 10.4, respectively. A fragment of SEQ ID NO:2 from about nucleotide 865 to about nucleotide 918 is useful in hybridization or amplification technologies to 30 identify SEQ ID NO:2 and to distinguish between SEQ ID NO:2 and a related sequence. Northern analysis shows the expression of this sequence in various libraries, at least 45% of which are associated with cancer, at least 20% of which are associated with developing and proliferating tissues or cells, and at least 25% of which are associated with the immune/inflammatory response. Of particular note is the expression of LRSP in brain and pituitary (30%), gut (25%), and reproductive

The invention also encompasses LRSP variants. A preferred LRSP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the LRSP amino acid sequence, and which contains at least one functional or structural characteristic of LRSP.

The invention also encompasses polynucleotides which encode LRSP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:2, which encodes LRSP.

The invention also encompasses a variant of a polynucleotide sequence encoding LRSP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding LRSP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:2 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:2. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of LRSP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding LRSP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring LRSP, and all such variations are to be considered as being specifically disclosed.

18

WO 00/24779

PCT/US99/24046

a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode LRSP may be cloned in recombinant DNA molecules that direct expression of LRSP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express LRSP.

The nucleotide sequences of the present invention can be engineered using methods generally

5 mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding LRSP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.)

Alternatively, LRSP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of LRSP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY.)

In order to express a biologically active LRSP, the nucleotide sequences encoding LRSP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding LRSP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding LRSP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding LRSP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of

WO 00/24779

PCT/US99/24046

enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding LRSP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding LRSP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding LRSP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding LRSP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding LRSP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of LRSP are needed, e.g. for the production of antibodies, vectors which direct high level expression of LRSP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of LRSP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

WO 00/24779

PCT/US99/24046

Plant systems may also be used for expression of LRSP. Transcription of sequences encoding LRSP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding LRSP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses LRSP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of LRSP in cell lines is preferred. For example, sequences encoding LRSP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et

al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g.,

5 Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β -glucuronide, or luciferase and its substrate
10 luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the
15 sequence encoding LRSP is inserted within a marker gene sequence, transformed cells containing sequences encoding LRSP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding LRSP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

20 In general, host cells that contain the nucleic acid sequence encoding LRSP and that express LRSP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

25 Immunological methods for detecting and measuring the expression of LRSP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on LRSP is preferred, but a
30 competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding LRSP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide.

5 Alternatively, the sequences encoding LRSP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega
10 (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding LRSP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein
15 produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode LRSP may be designed to contain signal sequences which direct secretion of LRSP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the
20 inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for
25 post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding LRSP may be ligated to a heterologous sequence resulting in translation of a
30 fusion protein in any of the aforementioned host systems. For example, a chimeric LRSP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of LRSP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST),

WO 00/24779

PCT/US99/24046

maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the LRSP encoding sequence and the heterologous protein sequence, so that LRSP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled LRSP may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

Fragments of LRSP may be produced not only by recombinant means, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, *supra*, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin-Elmer). Various fragments of LRSP may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of LRSP, and lysine-rich statherin protein and basic histidine-rich protein. In addition, the expression of LRSP is closely associated with cancer and the immune/inflammatory response and with disorders and conditions involving physiological stress. Therefore, LRSP appears to play a role in neoplastic and autoimmune/inflammatory disorders, and infectious diseases. In the treatment of disorders associated with increased LRSP expression or activity, it is desirable to decrease the expression or activity of LRSP. In the treatment of disorders associated with decreased LRSP expression or activity, it is desirable to increase the expression or activity of LRSP.

Therefore, in one embodiment, LRSP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LRSP. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress

syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and an infectious disease, such as infections by bacterial agents classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, and campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, and mycoplasma; and infections by fungal agents classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malassezia, histoplasma, and other fungal agents causing various mycoses.

In another embodiment, a vector capable of expressing LRSP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LRSP including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified LRSP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LRSP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of LRSP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LRSP including, but not limited to, those listed above.

In a further embodiment, an antagonist of LRSP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of LRSP. Examples of such disorders include, but are not limited to, a neoplastic disorder, such as, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin,

molecule may be produced.

Monoclonal antibodies to LRSP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce LRSP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for LRSP may also be generated. For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between LRSP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering LRSP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay

techniques may be used to assess the affinity of antibodies for LRSP. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of LRSP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions.

The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple LRSP epitopes, represents the average affinity, or avidity, of the antibodies for LRSP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular LRSP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the LRSP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of LRSP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J.E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of LRSP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding LRSP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding LRSP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding LRSP. Thus, complementary molecules or fragments may be used to modulate LRSP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding LRSP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding LRSP. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding LRSP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding LRSP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding LRSP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may be employed. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding LRSP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

15 Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of LRSP, antibodies to LRSP, and mimetics, agonists, antagonists, or inhibitors of LRSP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing

of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic

amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of LRSP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example LRSP or fragments thereof, antibodies of LRSP, and agonists, antagonists or inhibitors of LRSP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

In another embodiment, antibodies which specifically bind LRSP may be used for the diagnosis of disorders characterized by expression of LRSP, or in assays to monitor patients being treated with LRSP or agonists, antagonists, or inhibitors of LRSP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for LRSP include methods which utilize the antibody and a label to detect LRSP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

35

In another embodiment of the invention, the polynucleotides encoding LRSP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of LRSP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of LRSP, and to monitor regulation of LRSP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding LRSP or closely related molecules may be used to identify nucleic acid sequences which encode LRSP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding LRSP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the LRSP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:2 or from genomic sequences including promoters, enhancers, and introns of the LRSP gene.

Means for producing specific hybridization probes for DNAs encoding LRSP include the cloning of polynucleotide sequences encoding LRSP or LRSP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding LRSP may be used for the diagnosis of disorders associated with expression of LRSP. Examples of such disorders include, but are not limited to, a neoplastic disorder, such as, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes

mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and an infectious disease, such as infections by bacterial agents classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, and campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, and mycoplasma; and infections by fungal agents classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malassezia, histoplasma, and other fungal agents causing various mycoses. The polynucleotide sequences encoding LRSP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered LRSP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding LRSP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding LRSP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding LRSP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of LRSP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding LRSP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects

with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

5 Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

10 With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development
15 or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding LRSP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding LRSP, or a fragment of a polynucleotide complementary to the polynucleotide encoding
20 LRSP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

Methods which may also be used to quantify the expression of LRSP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from
25 standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

30 In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and

monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding LRSP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding LRSP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, LRSP, its catalytic or immunogenic fragments, or

oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between LRSP and the agent being tested may be measured.

5 Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with LRSP, or fragments thereof, and washed. Bound LRSP is then detected by methods well known in the art. Purified LRSP can also
10 be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding LRSP specifically compete with a test compound for binding LRSP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more
15 antigenic determinants with LRSP.

In additional embodiments, the nucleotide sequences which encode LRSP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

20 Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in
25 particular U.S. Ser. No. 60/155,209, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

The BRSTNOT14 library was constructed using RNA isolated from breast tissue removed
30 from a 62-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive grade 3 (of 4), nuclear grade 3 (of 3) adenocarcinoma. Patient history included a benign colon neoplasm, hyperlipidemia, cardiac dysrhythmia, and obesity. Family history included cardiovascular and cerebrovascular disease and colon, ovary and lung cancer.

The frozen tissue was homogenized and lysed in TRIZOL reagent (1 g tissue/10 ml TRIZOL; Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate, using a Brinkmann Homogenizer Polytron PT-3000 (Brinkmann Instruments, Westbury NY). After a brief incubation on ice, chloroform was added (1:5 v/v) and the lysate was centrifuged. The upper
5 chloroform layer was removed to a fresh tube and the RNA extracted with isopropanol, resuspended in DEPC-treated water, and treated with DNase for 25 min at 37°C.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN,
10 Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP
15 vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-
20 1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), or pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into competent *E. coli*
25 cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or
30 WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

cDNA sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Perkin-Elmer) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Perkin-Elmer) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 1 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 1 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

WO 00/24779

PCT/US99/24046

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:2. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower

scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding LRSP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in the description of the invention.

V. Extension of LRSP Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:2 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and β -mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar,

Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequences of SEQ ID NO:2 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:2 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ -³²P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston

MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10⁷ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, *supra*.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the LRSP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring LRSP. Although use of oligonucleotides

comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of LRSP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the LRSP-encoding transcript.

IX. Expression of LRSP

Expression and purification of LRSP is achieved using bacterial or virus-based expression systems. For expression of LRSP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).

Antibiotic resistant bacteria express LRSP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of LRSP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding LRSP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, LRSP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from LRSP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995,

supra, ch. 10 and 16). Purified LRSP obtained by these methods can be used directly in the following activity assay.

X. Demonstration of LRSP Activity

The activity of LRSP may be demonstrated by measuring the effect that LRSP has upon
5 bacterial growth. Antifungal and antibacterial agents inhibit bacterial cell growth (Mor, A. and P. Nicolas (1994) J. Biol. Chem. 269:1934-1939). Bacterial cells, such as *E. coli* (0.05 OD₆₀₀ units), are incubated with LB medium containing various amounts of LRSP (5 to 1000 µg/ml) for 24 hours at 30°C in 96-well plates as described by Mor and Nicolas (supra). The absorbance of the cells in each well is then measured at 600 nm or 490 nm using a 96-well plate reader. The absorbance of the cells
10 is compared to that of control, untreated cells. The decrease in absorbance is proportional to the amount of LRSP in the sample.

Alternatively, LRSP, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled LRSP,
15 washed, and any wells with labeled LRSP complex are assayed. Data obtained using different concentrations of LRSP are used to calculate values for the number, affinity, and association of LRSP with the candidate molecules.

XI. Functional Assays

LRSP function is assessed by expressing the sequences encoding LRSP at physiologically
20 elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 µg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome
25 formulations or electroporation. 1-2 µg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-
30 based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light

scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of LRSP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding LRSP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding LRSP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XII. Production of LRSP Specific Antibodies

LRSP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the LRSP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for anti-peptide and anti-LRSP activity by, for example, binding the peptide or LRSP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIII. Purification of Naturally Occurring LRSP Using Specific Antibodies

Naturally occurring or recombinant LRSP is substantially purified by immunoaffinity chromatography using antibodies specific for LRSP. An immunoaffinity column is constructed by covalently coupling anti-LRSP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is

blocked and washed according to the manufacturer's instructions.

Media containing LRSP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of LRSP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt

5 antibody/LRSP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and LRSP is collected.

XIV. Identification of Molecules Which Interact with LRSP

LRSP, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., Bolton and Hunter, supra.) Candidate molecules previously arrayed in the wells of a multi-
10 well plate are incubated with the labeled LRSP, washed, and any wells with labeled LRSP complex are assayed. Data obtained using different concentrations of LRSP are used to calculate values for the number, affinity, and association of LRSP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention.
15 Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater, fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res. 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger, and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score= 10-50 bits for PFAM hits, depending on individual protein families

Table 1 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Normalized quality score \geq GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. <i>supra</i> ; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof.

2. A substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of claim 1.

3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.

4. An isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide of claim 3.

5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.

6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 3.

7. A method for detecting a polynucleotide, the method comprising the steps of:
(a) hybridizing the polynucleotide of claim 6 to at least one nucleic acid in a sample, thereby forming a hybridization complex; and
(b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide in the sample.

8. The method of claim 7 further comprising amplifying the polynucleotide prior to hybridization.

9. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:2 and fragments thereof.

10. An isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide of claim 9.

11. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 9.
12. An expression vector comprising at least a fragment of the polynucleotide of claim 3.
13. A host cell comprising the expression vector of claim 12.
14. A method for producing a polypeptide, the method comprising the steps of:
- a) culturing the host cell of claim 13 under conditions suitable for the expression of the polypeptide; and
 - b) recovering the polypeptide from the host cell culture.
15. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.
16. A purified antibody which specifically binds to the polypeptide of claim 1.
17. A purified agonist of the polypeptide of claim 1.
18. A purified antagonist of the polypeptide of claim 1.
19. A method for treating or preventing a disorder associated with decreased expression or activity of LRSP, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 15.
20. A method for treating or preventing a disorder associated with increased expression or activity of LRSP, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 18.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

09030244 01123001



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C07K 14/81, 14/435, C12N 15/12		A1	(11) International Publication Number: WO 00/24779																																																																																																																																																																																																																																																																																																																																			
			(43) International Publication Date: 4 May 2000 (04.05.00)																																																																																																																																																																																																																																																																																																																																			
(21) International Application Number: PCT/US99/24046		(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).																																																																																																																																																																																																																																																																																																																																				
(22) International Filing Date: 22 October 1999 (22.10.99)																																																																																																																																																																																																																																																																																																																																						
(30) Priority Data: 60/155,209 23 October 1998 (23.10.98) US		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).																																																																																																																																																																																																																																																																																																																																				
(63) Related by Continuation (CON) of Continuation-in-Part (CIP) to Earlier Application US 60/155,209 (CIP) Filed on 23 October 1998 (23.10.98)		Published With international search report.																																																																																																																																																																																																																																																																																																																																				
(71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).																																																																																																																																																																																																																																																																																																																																						
(72) Inventors; and (75) Inventors/Applicants (for US only): TANG, Y., Tom [CH/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). CORLEY, Neil C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US).																																																																																																																																																																																																																																																																																																																																						
(54) Title: LYSINE-RICH STATHERIN PROTEIN																																																																																																																																																																																																																																																																																																																																						
<table><tr><td>5'</td><td>CTA</td><td>TGT</td><td>TTT</td><td>TAG</td><td>AAT</td><td>CAA</td><td>AGA</td><td>TGA</td><td>ACC</td><td>GGT</td><td>AAG</td><td>CTG</td><td>TCT</td><td>CAT</td><td>GTA</td><td>CCA</td><td>AAC</td><td>GTG</td></tr><tr><td></td><td></td><td></td><td>63</td><td></td><td></td><td>72</td><td></td><td></td><td>81</td><td></td><td></td><td>90</td><td></td><td></td><td>99</td><td></td><td></td><td>108</td></tr><tr><td></td><td>AAA</td><td>TTT</td><td>ACA</td><td>GTG</td><td>TTT</td><td>ACA</td><td>AAT</td><td>GTC</td><td>TGG</td><td>AAT</td><td>TTT</td><td>GCA</td><td>CTG</td><td>CCA</td><td>TAG</td><td>GGA</td><td>ATG</td><td>TTA</td></tr><tr><td></td><td></td><td>117</td><td></td><td></td><td>126</td><td></td><td></td><td>135</td><td></td><td></td><td>144</td><td></td><td></td><td>153</td><td></td><td></td><td>162</td><td></td></tr><tr><td></td><td>AGG</td><td>TTA</td><td>CTT</td><td>GGC</td><td>TGG</td><td>AAT</td><td>TTA</td><td>TCA</td><td>GAC</td><td>TTG</td><td>TGA</td><td>GTA</td><td>AAC</td><td>AAG</td><td>TTG</td><td>AAG</td><td>TTT</td><td>AGC</td></tr><tr><td></td><td></td><td>171</td><td></td><td></td><td>180</td><td></td><td></td><td>189</td><td></td><td></td><td>198</td><td></td><td></td><td>207</td><td></td><td></td><td>216</td><td></td></tr><tr><td></td><td>AGA</td><td>TGA</td><td>GGG</td><td>GGA</td><td>ATA</td><td>TTG</td><td>AGG</td><td>CCC</td><td>CTA</td><td>AGG</td><td>CTA</td><td>AAC</td><td>AAA</td><td>ATA</td><td>ATC</td><td>AGT</td><td>ATC</td><td>TGA</td></tr><tr><td></td><td></td><td>225</td><td></td><td></td><td>234</td><td></td><td></td><td>243</td><td></td><td></td><td>252</td><td></td><td></td><td>261</td><td></td><td></td><td>270</td><td></td></tr><tr><td></td><td>GAT</td><td>AGT</td><td>GGC</td><td>TAA</td><td>TGT</td><td>GGC</td><td>TCC</td><td>CCA</td><td>GGC</td><td>CTA</td><td>ATT</td><td>TGG</td><td>GAA</td><td>CAG</td><td>TTT</td><td>TTC</td><td>CTG</td><td>ATT</td></tr><tr><td></td><td></td><td>279</td><td></td><td></td><td>288</td><td></td><td></td><td>297</td><td></td><td></td><td>306</td><td></td><td></td><td>315</td><td></td><td></td><td>324</td><td></td></tr><tr><td></td><td>GCT</td><td>TTG</td><td>AGA</td><td>AGT</td><td>ACT</td><td>TTC</td><td>TTT</td><td>TGA</td><td>CAG</td><td>AAA</td><td>TTT</td><td>TCA</td><td>TTC</td><td>TGC</td><td>TTG</td><td>CCA</td><td>TTG</td><td>CTA</td></tr><tr><td></td><td></td><td>333</td><td></td><td></td><td>342</td><td></td><td></td><td>351</td><td></td><td></td><td>360</td><td></td><td></td><td>369</td><td></td><td></td><td>378</td><td></td></tr><tr><td></td><td>TAT</td><td>TCT</td><td>CCC</td><td>TTT</td><td>ATA</td><td>GGA</td><td>GCC</td><td>ATT</td><td>GGA</td><td>TTT</td><td>CTT</td><td>TCC</td><td>TTT</td><td>TGT</td><td>GGG</td><td>AAA</td><td>TGT</td><td>CCC</td></tr><tr><td></td><td></td><td>387</td><td></td><td></td><td>396</td><td></td><td></td><td>405</td><td></td><td></td><td>414</td><td></td><td></td><td>423</td><td></td><td></td><td>432</td><td></td></tr><tr><td></td><td>ATT</td><td>AGC</td><td>ATT</td><td>TTC</td><td>AGA</td><td>TCT</td><td>TTT</td><td>GAT</td><td>GTG</td><td>CAC</td><td>TAA</td><td>TGC</td><td>CAT</td><td>TAT</td><td>TGG</td><td>TAA</td><td>TGC</td><td>CGT</td></tr><tr><td></td><td></td><td>441</td><td></td><td></td><td>450</td><td></td><td></td><td>459</td><td></td><td></td><td>468</td><td></td><td></td><td>477</td><td></td><td></td><td>486</td><td></td></tr><tr><td></td><td>TAT</td><td>TGG</td><td>TGA</td><td>ATA</td><td>CAG</td><td>CAT</td><td>AGT</td><td>TAA</td><td>ATA</td><td>AAC</td><td>TGT</td><td>TAC</td><td>AGT</td><td>AAA</td><td>TCT</td><td>ACA</td><td>CTT</td><td>GGA</td></tr></table>				5'	CTA	TGT	TTT	TAG	AAT	CAA	AGA	TGA	ACC	GGT	AAG	CTG	TCT	CAT	GTA	CCA	AAC	GTG				63			72			81			90			99			108		AAA	TTT	ACA	GTG	TTT	ACA	AAT	GTC	TGG	AAT	TTT	GCA	CTG	CCA	TAG	GGA	ATG	TTA			117			126			135			144			153			162			AGG	TTA	CTT	GGC	TGG	AAT	TTA	TCA	GAC	TTG	TGA	GTA	AAC	AAG	TTG	AAG	TTT	AGC			171			180			189			198			207			216			AGA	TGA	GGG	GGA	ATA	TTG	AGG	CCC	CTA	AGG	CTA	AAC	AAA	ATA	ATC	AGT	ATC	TGA			225			234			243			252			261			270			GAT	AGT	GGC	TAA	TGT	GGC	TCC	CCA	GGC	CTA	ATT	TGG	GAA	CAG	TTT	TTC	CTG	ATT			279			288			297			306			315			324			GCT	TTG	AGA	AGT	ACT	TTC	TTT	TGA	CAG	AAA	TTT	TCA	TTC	TGC	TTG	CCA	TTG	CTA			333			342			351			360			369			378			TAT	TCT	CCC	TTT	ATA	GGA	GCC	ATT	GGA	TTT	CTT	TCC	TTT	TGT	GGG	AAA	TGT	CCC			387			396			405			414			423			432			ATT	AGC	ATT	TTC	AGA	TCT	TTT	GAT	GTG	CAC	TAA	TGC	CAT	TAT	TGG	TAA	TGC	CGT			441			450			459			468			477			486			TAT	TGG	TGA	ATA	CAG	CAT	AGT	TAA	ATA	AAC	TGT	TAC	AGT	AAA	TCT	ACA	CTT	GGA
5'	CTA	TGT	TTT	TAG	AAT	CAA	AGA	TGA	ACC	GGT	AAG	CTG	TCT	CAT	GTA	CCA	AAC	GTG																																																																																																																																																																																																																																																																																																																				
			63			72			81			90			99			108																																																																																																																																																																																																																																																																																																																				
	AAA	TTT	ACA	GTG	TTT	ACA	AAT	GTC	TGG	AAT	TTT	GCA	CTG	CCA	TAG	GGA	ATG	TTA																																																																																																																																																																																																																																																																																																																				
		117			126			135			144			153			162																																																																																																																																																																																																																																																																																																																					
	AGG	TTA	CTT	GGC	TGG	AAT	TTA	TCA	GAC	TTG	TGA	GTA	AAC	AAG	TTG	AAG	TTT	AGC																																																																																																																																																																																																																																																																																																																				
		171			180			189			198			207			216																																																																																																																																																																																																																																																																																																																					
	AGA	TGA	GGG	GGA	ATA	TTG	AGG	CCC	CTA	AGG	CTA	AAC	AAA	ATA	ATC	AGT	ATC	TGA																																																																																																																																																																																																																																																																																																																				
		225			234			243			252			261			270																																																																																																																																																																																																																																																																																																																					
	GAT	AGT	GGC	TAA	TGT	GGC	TCC	CCA	GGC	CTA	ATT	TGG	GAA	CAG	TTT	TTC	CTG	ATT																																																																																																																																																																																																																																																																																																																				
		279			288			297			306			315			324																																																																																																																																																																																																																																																																																																																					
	GCT	TTG	AGA	AGT	ACT	TTC	TTT	TGA	CAG	AAA	TTT	TCA	TTC	TGC	TTG	CCA	TTG	CTA																																																																																																																																																																																																																																																																																																																				
		333			342			351			360			369			378																																																																																																																																																																																																																																																																																																																					
	TAT	TCT	CCC	TTT	ATA	GGA	GCC	ATT	GGA	TTT	CTT	TCC	TTT	TGT	GGG	AAA	TGT	CCC																																																																																																																																																																																																																																																																																																																				
		387			396			405			414			423			432																																																																																																																																																																																																																																																																																																																					
	ATT	AGC	ATT	TTC	AGA	TCT	TTT	GAT	GTG	CAC	TAA	TGC	CAT	TAT	TGG	TAA	TGC	CGT																																																																																																																																																																																																																																																																																																																				
		441			450			459			468			477			486																																																																																																																																																																																																																																																																																																																					
	TAT	TGG	TGA	ATA	CAG	CAT	AGT	TAA	ATA	AAC	TGT	TAC	AGT	AAA	TCT	ACA	CTT	GGA																																																																																																																																																																																																																																																																																																																				
(57) Abstract																																																																																																																																																																																																																																																																																																																																						
<p>The invention provides a human lysine-rich statherin protein (LRSP) and polynucleotides which identify and encode LRSP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of LRSP.</p>																																																																																																																																																																																																																																																																																																																																						

5' CTA TGT TTT TAG AAT CAA AGA TGA ACC GGT AAG CTG TCT CAT GTA CCA AAC GTG
 9 18 27 36 45 54
 63 72 81 90 99 108
 AAA TTT ACA GTG TTT ACA AAT GTC TGG AAT TTT GCA CTG CCA TAG GGA ATG TTA
 117 126 135 144 153 162
 AGG TTA CTT GGC TGG AAT TTA TCA GAC TTG TGA GTA AAC AAG TTG AAG TTT AGC
 171 180 189 198 207 216
 AGA TGA GCG GGA ATA TTG AGG CCC CTA AGG CTA AAC AAA ATA ATC AGT ATC TGA
 225 234 243 252 261 270
 GAT AGT GGC TAA TGT GGC TCC CCA GGC CTA ATT TGG GAA CAG TTT TTC CTG ATT
 279 288 297 306 315 324
 GCT TTG AGA AGT ACT TTC TTT TGA CAG AAA TTT TCA TTC TGC TTG CCA TTG CTA
 333 342 351 360 369 378
 TAT TCT CCC TTT ATA GGA GCC ATT GGA TTT CTT TCC TTT TGT GGG AAA TGT CCC
 387 396 405 414 423 432
 ATT AGC ATT TTC AGA TCT TTT GAT GTG CAC TAA TGC CAT TAT TGG TAA TGC CGT
 441 450 459 468 477 486
 TAT TGG TGA ATA CAG CAT AGT TAA ATA AAC TGT TAC AGT AAA TCT ACA CTT GGA

FIGURE 1A

495 504 513 522 531 540
 TTT GCT GCA CCT CTA CCA ATA GCC TTT TGA ATG ACT GAA AGT GTT AAC AGA GAA

 549 558 567 576 585 594
 AGA GGC ATG TCT GCA GAA AGA GAT AGC TAA TAT TTT TTG GTA CTT TAT CTG AAA

 603 612 621 630 639 648
 TCC AAG ATG CTG CTT CCC CTG CAG GTT GTT TTC CTT CTT ACG ATC CTC ATT GAA

 657 666 675 684 693 702
 TCC CCT CTG GGA GCA CAG GAC AGT TAG TAG AAC TCT CCA TTT CTT TGT TTT GTT

 711 720 729 738 747 756
 TTT TAA GAC AGA GAC TCT GTC TCA AAA AAA AGG ACA TTT ATC ATT ATA ACA TCT

 765 774 783 792 801 810
 TAT TAG AGC CCC TAA TTT CTT ATC TGA AGG CAC TGT TTT TTT TAA ACA GTT

 819 828 837 846 855 864
 AAG TAC TGA TGT CAA CAG ACA AAT ATT TCT GAT CAG ATA GTC CCC TGT CAA CAG

 873 882 891 900 909 918
 TAG CAA ATG TGG TTT CAT AAA GTG GGA AGA AAA CAG CAT TTT AAA GTA ACT TTT
 M W F H K V G R K Q H F K V T F

 927 936 945 954 963 972
 TGG GAG ACT GAT TTG AGT AAT AAT AAA ACT CTG GTC TCC CTT AAG AAA AAA AAA
 W E T D L S N N K T L V S L K K K K

FIGURE 1B

981 1035 1089 1143 1197 1251 1305
CCC TTC CAC CTT TAC TGT GTC ATT TAT ATC CCC TTA GTT CCA AAG TTA ATT ATC
P F H L Y C V I Y I I P L V P K L I I
TTA TTT CTG GAT ATT GCT TTT ATA CCA AAG AGC CTT ATC AGC CAG TTC CAG AAC
L F L D I A F I P K S L I S Q F Q N
AAC CAC TAT ACG CAC AAC CAT ACC AAC CAC AAT ACC AAC AAT ATA CGT TTT AAT
N H Y T H N H T N H N T N N I R F N
ATC ATC AGT AAC TGC AGG ACA TGA TTA TTG AGG CTT GAT TGG CAA ATA CGA CTT
I I S N C R T
CTA CAT CCA TAT TCT CAT CTT TCA TAC CAT ATC ACA CTA CTA CCA CTT TTT GTN
AGA TCA TCT AAG AGC AAT GCG AAT GTA AAA CCC TAT AAT TTA CTG GAT ACT CTT
TGG TTC CAG ATA CTT GCC TTT TCC AAT GTC ACT TG 3'

FIGURE 1C

1	C	T	A	T	G	T	T	T	T	A	G	A	A	T	C	A	A	G	A	T	G	A	A	C	C	G	G	T	2820214	
1	A	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 338507	
1	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 179465	
31	A	A	G	C	T	G	T	C	T	C	A	T	G	T	A	C	C	A	A	A	C	G	T	G	A	A	T	T	2820214	
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 338507	
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 179465	
61	A	C	A	G	T	G	T	T	T	A	C	A	A	T	G	T	C	T	G	G	A	A	T	T	T	T	G	C	A	2820214
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 338507	
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 179465	
91	C	T	G	C	C	A	T	A	G	G	A	A	T	G	T	T	A	A	G	G	T	T	A	C	T	T	G	G	C	2820214
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 338507	
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 179465	
121	T	G	G	A	A	T	T	A	T	C	A	G	A	C	T	T	G	T	G	A	G	T	A	A	A	C	A	A	G	2820214
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 338507	
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 179465	
151	T	T	G	A	A	G	T	T	T	A	G	C	A	G	A	T	G	A	G	G	G	G	A	T	A	T	T	G	2820214	
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 338507	
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 179465	

FIGURE 2A

181	AGGCCCTAAGGCTAAACAATAATCAGT																											2820214
3	--CCTTGAGCTTCAC--																											GI 338507
2	-----																											GI 179465
211	ATCTGAGATAGTGGCTAATGTGGCTCCCCA																											2820214
18	-----																											GI 338507
2	-----																											GI 179465
241	GGCCTAATTGGGAACAGTTTTCCTGATT																											2820214
18	-----																											GI 338507
2	-----																											GI 179465
271	GCTTTGAGAAAGTACTTTCTTTTGACAGAAA																											2820214
18	-----																											GI 338507
2	-----																											GI 179465
301	TTTTCATTCTGCTTGCCATTGCTATATCT																											2820214
18	-----																											GI 338507
2	-----																											GI 179465
331	CCCTTTATAGGAGCCATTGGATTTCCTTCC																											2820214
18	--TTCAACCTTCACTAC																											GI 338507
2	-----																											GI 179465

FIGURE 2B

09630244_042301

361	T T T T G T G G G A A A T G T C C C A T T A G C A T T T C	2820214
32	T T C T G T -	GI 338507
2	- -	GI 179465
391	A G A T C T T T T G A T G T G C A C T A A T G C C A T T A T	2820214
38	- -	GI 338507
2	- -	GI 179465
421	T G G T A A T G C C G T T A T T G G T G A A T A C A G C A T	2820214
38	- -	GI 338507
2	- -	GI 179465
451	A G T T A A A T A A A C T G T T A C A G T A A A T C T A C A	2820214
38	- -	GI 338507
2	- -	GI 179465
481	C T T G G A T T T G C T G C C A C C T C T A C C A A T A G C C	2820214
38	- -	GI 338507
2	- -	GI 179465
511	T T T T G A A T G A C T G A A A G T G T T A A C A G A G A A	2820214
38	- -	GI 338507
2	- -	GI 179465

FIGURE 2C

FIGURE 2D

721	G T C T C A A A A A A G G A C A T T T A T C A T T A T A																									2820214
120	- -																									GI 338507
85	- -																									GI 179465
751	A C A T C T T A T T A G A G C C C C T A A T T T C T T A T C																									2820214
120	- - - - - G A T T G G A G C - - - - - T G A T T C A T C																									GI 338507
85	- - - - - G A C T G G A G C - - - - - T G A T T C A C A																									GI 179465
781	T G A A G G C A C T G T T T T T T T T A A A C A G T T																									2820214
138	T G A A G -																									GI 338507
103	T G C A A -																									GI 179465
811	A A G T A C T G A T G T C A A C A G A C A A A T A T T T C T																									2820214
143	- - - - - - - - - - - - - - - - A G A A T T T T T																									GI 338507
108	- - - - - - - - - - - - - - - - A G A G A C A T C -																									GI 179465
841	G A T C A G A T A G T C C C C T G T C A A C A G T A G C A A																									2820214
153	G -																									GI 338507
117	- -																									GI 179465
871	A T G T G G T T T C A T A A A G T G G G A A G A A A C A G																									2820214
154	- - - - - - - - - - C G T A G A A T T G G A A G A - - - - -																									GI 338507
117	- - - - - - - - - - - - - - - A T G G G T A T A A A G A A																									GI 179465

FIGURE 2E

901	C	A	T	T	T	A	A	G	T	A	A	C	T	T	T	T	T	G	G	G	A	G	A	C	T	G	A	T	2820214		
169	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 338507			
132	A	A	T	T	C	C	A	T	G	A	A	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 179465			
931	T	T	G	A	G	T	A	A	T	A	A	A	A	C	T	C	T	G	G	T	C	T	C	C	C	T	T	2820214			
169	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 338507			
145	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 179465			
961	A	A	G	A	A	A	A	A	A	A	A	C	C	C	T	T	C	C	A	C	C	T	T	T	A	C	T	G	T	2820214	
169	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 338507			
145	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 179465			
991	G	T	C	A	T	T	A	T	A	T	C	C	C	C	T	T	A	G	T	T	C	C	A	A	A	G	T	T	A	2820214	
169	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 338507			
147	A	T	C	A	T	T	C	A	C	A	T	C	G	A	-	-	G	G	C	T	A	T	A	G	A	T	C	A	A	GI 179465	
1021	A	T	T	A	T	C	T	T	A	T	T	T	C	T	G	G	A	T	A	T	T	G	C	T	T	T	T	A	T	A	2820214
169	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 338507			
174	A	T	T	A	T	C	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 179465			
1051	C	C	A	A	G	A	G	C	C	T	T	A	T	C	A	G	C	C	A	G	T	T	C	C	A	G	A	A	C	2820214	
187	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 338507			
181	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 179465			

FIGURE 2F

FIGURE 2G

1260	T G C G A A T G T A A A - - - A C C C T A T A A T T T A C T	2820214
388	T G C A A A T G A A A A - - - A C A C T A T A A T T T A C T	GI 338507
321	T G C A G A A T A A A A G A A A T A C C A T G A T T T A G T	GI 179465
1287	G G A T A C T C T T T G G T T C C A G A T A C T T T G C C T T	2820214
415	G T A T A C T C T T T G T T T C A G G A T A C T T T G C C T T	GI 338507
351	G A A T - - T C T G T G T T T C A G G A T A C T T C C C T T	GI 179465
1317	T T C C A A T G T C A C T T G	2820214
445	T T C A A T T G T C A C T T G A T G A T A T A A T T G C A A	GI 338507
379	C C T A A T T A T C A T T T G A T T A G A T A C T T G C A A	GI 179465
1331	T T T A A A C T G T T A A G C T G T T C A G T A C T G T	2820214
475	T T T A A A - T G T T A A G C T G T T T T C A C T G C T G T	GI 338507
409		GI 179465
1331	T T C T G A A T A A T A G A A A T C A - - - C T T C T C T A	2820214
505	T T C T G A G T A A T A G A A A T T C A T T C C T C T C C A	GI 338507
438		GI 179465
1331	A A A G C A A T A A A T T T C A A G C C C	2820214
532	A A A G C A A T A A A T T T C A A G C A C A T T	GI 338507
468		GI 179465

FIGURE 2H

1	M	W	F	H	K	V	G	R	K	Q	H	F	K	V	T	F	W	E	T	D	L	S	N	N	K	T	L	V	S	L	K	K	K	P	2820214	
1	M	K	F	L	V	F	A	F	I	L	A	L	M	V	S	M	I	G	A	D	S	S	E	E	K	F	L	-	-	-	-	-	-	R	R	GI 338508
1	M	K	F	F	V	F	A	L	I	L	A	L	M	L	S	M	T	G	A	D	S	H	A	K	R	-	-	-	-	-	-	-	-	-	GI 179466	
36	F	H	L	Y	C	V	I	Y	I	P	L	V	P	K	L	I	I	L	F	L	D	I	A	F	I	P	K	S	L	I	S	Q	F	Q	N	2820214
30	I	G	R	F	G	Y	G	Y	G	P	Y	Q	P	-	-	-	-	-	V	P	E	Q	P	L	Y	P	Q	P	Y	Q	P	Q	Y	Q	-	GI 338508
26	-	-	H	H	G	Y	K	-	-	-	-	-	R	K	-	-	-	-	F	H	E	K	H	H	S	H	R	G	Y	R	S	N	Y	L	-	GI 179466
71	N	H	Y	T	H	N	H	T	N	H	N	T	N	N	I	R	F	N	I	I	S	N	C	R	T											2820214
59	-	Q	Y	T	F																														GI 338508	
49	-	-	Y	D	N																														GI 179466	

FIGURE 3

Docket No.: PF-0620 USN

**DECLARATION AND POWER OF ATTORNEY FOR
UNITED STATES PATENT APPLICATION**

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,
and

I believe that I am the original, first and sole inventor (if only one name is listed below)
or an original, first and joint inventor (if more than one name is listed below) of the subject
matter which is claimed and for which a United States patent is sought on the invention entitled

LYSINE-RICH STATHERIN PROTEIN

the specification of which:

 / is attached hereto.

 / was filed on _____ as application Serial No. _____ and if this box
contains an X /, was amended on _____.

 / was filed as Patent Cooperation Treaty international application No. PCT/US99/24046 on
October 22, 1999, if this box contains an X /, was amended on under Patent Cooperation
Treaty Article 19 on _____ 2001, and if this box contains an X /, was amended on _____.

I hereby state that I have reviewed and understand the contents of the above-identified
specification, including the claims, as amended by any amendment referred to above.

I acknowledge my duty to disclose information which is material to the examination of
this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim the benefit under Title 35, United States Code, §119 or §365(a)-(b) of any
foreign application(s) for patent or inventor's certificate indicated below and of any Patent
Cooperation Treaty international applications(s) designating at least one country other than the
United States indicated below and have also identified below any foreign application(s) for patent
or inventor's certificate and Patent Cooperation Treaty international application(s) designating at
least one country other than the United States for the same subject matter and having a filing date
before that of the application for said subject matter the priority of which is claimed:

Docket No.: PF-0620 USN

Country	Number	Filing Date	Priority Claimed
_____	_____	_____	// Yes // No
_____	_____	_____	// Yes // No

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)
60/155,209	October 22, 1999	Expired

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of Title 35, United States Code §112, I acknowledge my duty to disclose material information as defined in Title 37 Code of Federal Regulations, §1.56(a) which occurred between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)
_____	_____	_____

I hereby appoint the following:

Lucy J. Billings	Reg. No. <u>36,749</u>
Michael C. Cerrone	Reg. No. <u>39,132</u>
Diana Hamlet-Cox	Reg. No. <u>33,302</u>
Richard C. Ekstrom	Reg. No. <u>37,027</u>
Barrie D. Greene	Reg. No. <u>46,740</u>
Matthew R. Kaser	Reg. No. <u>44,817</u>
Lynn E. Murry	Reg. No. <u>42,918</u>
Shirley A. Recipon	Reg. No. <u>47,016</u>
Susan K. Sather	Reg. No. <u>44,316</u>
Michelle M. Stempien	Reg. No. <u>41,327</u>
David G. Streeter	Reg. No. <u>43,168</u>
Stephen Todd	Reg. No. <u>47,139</u>
Christopher Turner	Reg. No. <u>45,167</u>
P. Ben Wang	Reg. No. <u>41,420</u>

14

respectively and individually, as my patent attorneys and/or agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith. Please address all communications to:

Docket No.: PF-0620 USN

**LEGAL DEPARTMENT
 INCYTE GENOMICS, INC.
 3160 PORTER DRIVE, PALO ALTO, CA 94304**

TEL: 650-855-0555 FAX: 650-849-8886 or 650-845-4166

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

**Sole Inventor or
 First Joint Inventor:**

1-00

Full name:Y. TOM TANG**Signature:**Y. Tom Tang**Date:**February 27, 2001**Citizenship**People's Republic of China ^{USA} 02/27/2001**Residence:**San Jose, California CA**P.O. Address:**4230 Ranwick Court
 San Jose, California 95118

Second Joint Inventor:

2nd

Full name:NEIL C. CORLEY**Signature:**Neil C. Corley**Date:**MARCH 5, 2001**Citizenship**United States of America**Residence:**Castro Valley, California CA**P.O. Address:**20426 Crow Creek Road
 Castro Valley, California 94552

Docket No.: PF-0620 USN

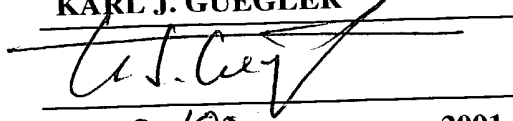
Third Joint Inventor:

3w

Full name:

KARL J. GUEGLER

Signature:



Date:

02/02, 2001

Citizenship

Switzerland

Residence:

Menlo Park, California CA

P.O. Address:

1048 Oakland Avenue
Menlo Park, California 94025

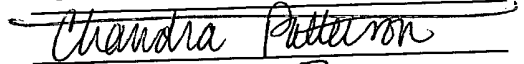
Fourth Joint Inventor:

4w

Full name:

CHANDRA PATTERSON

Signature:



Date:

February 7, 2001

Citizenship

United States of America

Residence:

Menlo Park, California CA

P.O. Address:

490 Sherwood Way
Menlo Park, California 94025

<110> INCYTE PHARMACEUTICALS, INC.

TANG, Y. Tom

CORLEY, Neil C.

GUEGLER, Karl J.

PATTERSON, Chandra

<120> LYSINE-RICH STATHERIN PROTEIN

<130> PF-0620 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/155,209

<151> 1998-10-23

<160> 6

<170> PERL Program

<210> 1

<211> 95

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2820214CD1

<400> 1

Met	Trp	Phe	His	Lys	Val	Gly	Arg	Lys	Gln	His	Phe	Lys	Val	Thr
1				5					10					15
Phe	Trp	Glu	Thr	Asp	Leu	Ser	Asn	Asn	Lys	Thr	Leu	Val	Ser	Leu
				20					25					30
Lys	Lys	Lys	Lys	Pro	Phe	His	Leu	Tyr	Cys	Val	Ile	Tyr	Ile	Pro
				35					40					45
Leu	Val	Pro	Lys	Leu	Ile	Ile	Leu	Phe	Leu	Asp	Ile	Ala	Phe	Ile
				50					55					60
Pro	Lys	Ser	Leu	Ile	Ser	Gln	Phe	Gln	Asn	Asn	His	Tyr	Thr	His
				65					70					75
Asn	His	Thr	Asn	His	Asn	Thr	Asn	Asn	Ile	Arg	Phe	Asn	Ile	Ile
				80					85					90
Ser	Asn	Cys	Arg	Thr										
				95										

<210> 2

<211> 1331

<212> DNA

<213> Homo sapiens

<210> 4
 <211> 51
 <212> PRT
 <213> Homo sapiens

<300> misc_feature
 <308> GenBank ID No: g179466

<400> 4
 Met Lys Phe Phe Val Phe Ala Leu Ile Leu Ala Leu Met Leu Ser
 1 5 10 15
 Met Thr Gly Ala Asp Ser His Ala Lys Arg His His Gly Tyr Lys
 20 25 30
 Arg Lys Phe His Glu Lys His His Ser His Arg Gly Tyr Arg Ser
 35 40 45
 Asn Tyr Leu Tyr Asp Asn
 50

<210> 5
 <211> 552
 <212> DNA
 <213> Homo sapiens

<300> misc_feature
 <308> GenBank ID No: g338507

<400> 5
 atctcttgaa gcttcacttc aacttcacta cttctgtagt ctcactcttga gtaaaagaga 60
 acccagccaa ctatgaagtt ccttgtcttt gccttcatct tggctctcat gggttccatg 120
 attggagctg attcatctga agagaaattt ttgcgtagaa ttggaagatt cggttatggg 180
 tatggccctt atcagccagt tccagaacaa ccactatacc cacaaccata ccaaccacaa 240
 taccaacaat atacctttta atatcatcag taactgcagg acatgattat tgaggcttga 300
 ttggcaataa cgacttctac atccatattc tcactcttca taccatatca cactactacc 360
 actttttgaa gaatcatcaa agagcaatgc aaatgaaaaa cactataatt tactgtatac 420
 tctttgtttc aggatacttg ctttttcaat tgtcacttga tgatataatt gcaatttaaa 480
 ctgttaagct gtgttcagta ctgtttctga ataatagaaa tcacttctct aaaagcaata 540
 aatttcaagc cc 552

<210> 6
 <211> 491
 <212> DNA
 <213> Homo sapiens

09830209/890244#8
 Rec'd PCT/PTO 13 DEC 2002

PF-0620 USN

<110> TANG, Y. Tom
 CORLEY, Neil C.
 GUEGLER, Karl J.
 PATTERSON, Chandra

<120> LYSINE-RICH STATHERIN PROTEIN

<130> PF-0620 USN

<140> 09/830,244

<141> Herewith

<150> PCT/US99/24046

<151> 1999-10-22

<150> 60/155,209

<151> 1998-10-23

<160> 6

<170> PERL Program

<210> 1

<211> 95

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2820214CD1

<400> 1

Met	Trp	Phe	His	Lys	Val	Gly	Arg	Lys	Gln	His	Phe	Lys	Val	Thr
1				5					10					15
Phe	Trp	Glu	Thr	Asp	Leu	Ser	Asn	Asn	Lys	Thr	Leu	Val	Ser	Leu
				20					25					30
Lys	Lys	Lys	Lys	Pro	Phe	His	Leu	Tyr	Cys	Val	Ile	Tyr	Ile	Pro
				35					40					45
Leu	Val	Pro	Lys	Leu	Ile	Ile	Leu	Phe	Leu	Asp	Ile	Ala	Phe	Ile
				50					55					60
Pro	Lys	Ser	Leu	Ile	Ser	Gln	Phe	Gln	Asn	Asn	His	Tyr	Thr	His
				65					70					75
Asn	His	Thr	Asn	His	Asn	Thr	Asn	Asn	Ile	Arg	Phe	Asn	Ile	Ile
				80					85					90
Ser	Asn	Cys	Arg	Thr										
				95										

<210> 2

<211> 1331

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2820214CB1

PF-0620 USN

<220>

<221> unsure

<222> 1242

<223> a, t, c, g, or other

<400> 2

```
ctatgtttttt agaatcaaag atgaaccggt aagctgtctc atgtaccaa cgtgaaattt 60
acagtgtttta caaatgtctg gaattttgca ctgccatagg gaatgttaag gttacttggc 120
tggaattttat cagacttggt agtaaacaag ttgaagttaa gcagatgagg gggaatattg 180
aggccccctaa ggctaaacaa aataatcagt atctgagata gtgggctaag tggctcccca 240
ggcctaattt gggaacagtt ttctctgatt gctttgagaa gtactttctt ttgacagaaa 300
ttttcattct gcttgccatt gctatattct ccttttatag gagccattgg atttctttcc 360
ttttgtggga aatgtcccat tagcattttc agatcttttg atgtgcacta atgccattat 420
tggtaatgcc gttattgggt aatacagcat agttaaataa actgttacag taaatctaca 480
cttggaattt ctgcacctct accaatagcc ttttgaatga ctgaaagtgt taacagagaa 540
agaggcatgt ctgcagaaag agatagctaa tatttttttg tactttatct gaaatccaag 600
atgtgtcttc cctgcaggt tgttttctt cttacgatcc tcattgaatc ccctctggga 660
gcacaggaca gttagtagaa ctctccattt ctttgttttg ttttttaaga cagagactct 720
gtctcaaaaa aaaggacatt tatcattata acatcttatt agagccccta atttcttctc 780
tgaaggcact gttttttttt ttaaacagtt aagtactgat gtcaacagac aaatatttct 840
gatcagatag tccccgttca acagtagcaa atgtgggttc ataaagtggg aagaaaacag 900
catttttaaag taactttttg ggagactgat ttgagtaata ataaaactct ggtctccctt 960
aagaaaaaaa aacccttcca cctttactgt gtcatttata tccccctagt tccaaagtta 1020
attatcttat ttctggatat tgcttttata ccaaagagcc ttatcagcca gttccagaac 1080
aaccactata cgcacaacca taccaaccac aataccaaca atatacgttt taatatcatc 1140
agtaactgca ggacatgatt attgaggctt gattggcaaa tacgacttct acatccatat 1200
tctcatcttt cataccatat cacactacta ccactttttg tnagatcatc taagagcaat 1260
ggaatgtaa aaccctataa ttactggat actcttttgg tccagatact tgccttttcc 1320
aatgtcactt g 1331
```

<210> 3

<211> 62

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> GenBank ID No: g338508

<400> 3

```
Met Lys Phe Leu Val Phe Ala Phe Ile Leu Ala Leu Met Val Ser
1 5 10 15
Met Ile Gly Ala Asp Ser Ser Glu Glu Lys Phe Leu Arg Arg Ile
20 25 30
Gly Arg Phe Gly Tyr Gly Tyr Gly Pro Tyr Gln Pro Val Pro Glu
35 40 45
Gln Pro Leu Tyr Pro Gln Pro Tyr Gln Pro Gln Tyr Gln Gln Tyr
50 55 60
Thr Phe
```

<210> 4

<211> 51

<212> PRT

[illegible]

<220>

<223> GenBank ID No: g179466

Met Lys Phe Phe Val Phe Ala Leu Ile Leu Ala Leu Met Leu Ser

Met Thr Gly Ala Asp Ser His Ala Lys Arg His His Gly Tyr Lys

Arg Lys Phe His Glu Lys His His Ser His Arg Gly Tyr Arg Ser

Asn Tyr Leu Tyr Asp Asn

<210> 5

<212> DNA

 $\langle 220 \rangle$

<223> GenBank ID No: g338507

atctcttgaa gcttcacttc aacttcacta cttctgtagt ctcatcttga gtaaaagaga 60

accagccaa ctatgaagtt ccttgtcttt gccttcatct tggctctcat ggtttccatg 120

attggagctg attcatctga agagaaattt ttgcgtagaa ttggaagatt cggttatggg 180

tatggccctt atcagccagt tccagaacaa ccactatacc cacaaccata ccaaccacaa 240

taccaacaat atacctttta atatcatcag taactgcagg acatgattat tgaggcttga 300

ttggcgaata cgacttctac atccatatcc tcattctttca taccatatca cactactacc 360

acttttttgaa gaatcatcaa agagcaatgc aaatgaaaaa cactataatt tactgtatac 420

tcttttgtttc aggatacttg ccttttcaat tgtcacttga tgatataatt gcaatttaaa 480

ctgttaagct gtgttcagta ctgtttctga ataataaaaa tcacttctct aaaagcaata 540

aattttcaagc cc 552

<210> 6

<211> 491

<212> DNA

 $\langle 220 \rangle$

```
<221> misc_feature
```

<223> GenBank ID No: g179465

<400> 6

aggacgccta ccaggaggac ctgggattca accaactatg aagttttttg tttttgcttt 60

aatctttggct ctcatgcttt ccatgactgg agctgattca catgcaaaga gacatcatgg 120

gtataaaaga aaattccatg aaaagcatca ttcacatcga ggctatagat caaattatct 180

gtatgacaat tgatatcttc agtaatcatg gggcatgatt atggaggttt gactggcaaa 240

ttcgccttctgg actcgtgtat tctcatttgt cataccgcac cacactacta ctgcttttttg 300

aagggaattat cataaggcaa tgcagaataa aagaaatacc atgatttagt gaattctgtg 360

tttcaggata cttcccttcc taattatcat ttgattagat acttgcaatt taaatggttaa 420

'PF-0620 USN

gctgttttca ctgctgtttc tgagtaatag aaattcattc ctctccaaaa gcaataaaat 480
tcaagcacat t 491